

Optimal Symmetric Multimodal Templates and Concatenated Random Forests for Supervised Brain Tumor Segmentation (Simplified) with ANTsR

Nicholas J. Tustison · K. L. Shrinidhi · Max Wintermark · Christopher R. Durst · Benjamin M. Kandel · James C. Gee · Murray C. Grossman · Brian B. Avants

Published online: 30 November 2014
© Springer Science+Business Media New York 2014

Abstract Segmenting and quantifying gliomas from MRI is an important task for diagnosis, planning intervention, and for tracking tumor changes over time. However, this task is complicated by the lack of prior knowledge concerning tumor location, spatial extent, shape, possible displacement of normal tissue, and intensity signature. To accommodate such complications, we introduce a framework for supervised segmentation based on multiple modality intensity, geometry, and asymmetry feature sets. These features drive a supervised whole-brain and tumor segmentation approach based on random forest-derived probabilities. The asymmetry-related features (based on optimal symmetric multimodal templates) demonstrate excellent discriminative properties within this framework. We also gain performance by generating probability maps from random forest models and using these maps for a refining Markov random field regularized probabilistic segmentation. This strategy allows us to interface the supervised learning capabilities of the random forest model with regularized probabilistic segmentation using the recently developed *ANTsR* package—a comprehensive statistical and visualization interface between the popular Advanced Normalization Tools (ANTs) and the *R* statistical project. The reported algorithmic framework was the top-performing entry in

the MICCAI 2013 Multimodal Brain Tumor Segmentation challenge. The challenge data were widely varying consisting of both high-grade and low-grade glioma tumor four-modality MRI from five different institutions. Average Dice overlap measures for the final algorithmic assessment were 0.87, 0.78, and 0.74 for “complete”, “core”, and “enhanced” tumor components, respectively.

Keywords Advanced normalization tools · BRATS · Glioma · Brain tumor segmentation · R project for statistical computing

Introduction

Given the complexity of tumor growth and appearance and the need for precise volumetric measurements for tumor characterization, much research has been invested in computational methods for automated segmentation of tumor regions in MR images. Several approaches have been previously proposed in the literature as detailed in recent reviews (Angelini et al. 2007; Bauer et al. 2013). Commensurable algorithmic evaluation, however, is extremely problematic due to widely varying performance assessments described in the corresponding publications, lack of publicly available evaluation data, and private algorithmic instantiations. In response to these issues, the Multimodal Brain Tumor Segmentation (BRATS) challenge was initiated in 2012 (and continued in 2013) under the auspices of the Medical Image Computing and Computer Assisted Intervention Society in association with their annual international conference (Menze et al. 2014).

After observing the 2012 challenge to gain insight for our own research needs (Durst et al. 2014), we found few

N. J. Tustison (✉) · M. Wintermark · C. R. Durst
Department of Radiology and Medical Imaging, University
of Virginia, Charlottesville, VA, USA
e-mail: ntustison@virginia.edu

K. L. Shrinidhi · B. M. Kandel · J. C. Gee ·
M. C. Grossman · B. B. Avants
Penn Image Computing and Science Laboratory, Department
of Radiology, University of Pennsylvania,
Philadelphia, PA, USA

implementation resources available. This prompted our participation in the challenge which took place the following year in Nagoya, Japan concerning brain tumor segmentation of a varied data set from five different institutions involving FLAIR, T1-weighted, T1-weighted contrast (T1C), and T2-weighted MRI modalities. Heavily influenced by the success of participants in the 2012 challenge (specifically, the work of Bauer et al. (2012), Geremia et al. (2012), and Zikic et al. (2012)), we adopted the increasingly popular random forest (RF) machine learning framework (Breiman 2001) which permits the inclusion of many potentially discriminative image features. Supervised segmentation techniques generally consist of a training phase for model construction using image-based feature data followed by prediction using the generated model. For supervised brain tumor segmentation, a set of training data consisting of labeled brain images is used to construct a predictive model. Although other classification techniques have been used to segment brain tumors (e.g., support vector machines (Bauer et al. 2011)), RF models have proven particularly successful.

Several machine learning concepts were integrated to create the RF framework first articulated in its entirety by Breiman (2001) for performing classification/regression. Although decision trees had been previously explored in the literature, it was the success of “boosting”-style machine learning techniques, such as AdaBoost (Schapire 1990; Freund and Schapire 1997), which influenced the aggregation of such decision trees into “forests” with randomized node optimization for improved classification/regression performance (Ho 1995; Amit and Geman 1997). The final element of bootstrap aggregating or “bagging” (i.e. random sampling of the training data) was introduced by Breiman (1996) to achieve improved accuracy.

Early adoption (Viola et al. 2005) and success in the computer vision community has led to a recent surge within the medical image analysis community of using RFs for handling complex classification/regression tasks including normal brain segmentation (Yi et al. 2009), MS lesion segmentation (Geremia et al. 2011), multimodal brain tumor segmentation (Bauer et al. 2012; Zikic et al. 2012), brain extraction (Iglesias et al. 2010), classification of Alzheimer’s disease (Gray et al. 2013), anatomy detection in computed tomography (Criminisi et al. 2013), and segmentation of echocardiographic images (Verhoek et al. 2011). A thorough introduction for those interested in delving deeper into the more theoretical aspects of RFs can be found in (Criminisi et al. 2011).

Instrumental to our success in the 2013 challenge was the use of image features based on multimodal shape and appearance asymmetries which were generated via the construction of symmetric multivariate brain templates. Addi-

tional improvements in performance were due to the use of a two-stage RF model approach whereby the output probability images from the application of the first RF model seeds the construction of a subset of new feature images for input into the second stage. The two-stage approach is motivated by the fact that the RF is a powerful tool for ranking the features in a high-dimensional model i.e., it can be used for model selection in voxel-wise classification. However, the RF outputs are not geometrically constrained, which is the motivation for our second refinement step.

In line with our philosophy regarding open science (Tustison et al. 2013; Ince et al. 2012), we have made the code available as open source and, to our knowledge, we are the only competitors out of the approximately 20 teams that participated in either years’ challenges to do this. All code is based on the well-vetted and open source Insight Toolkit (ITK) of the National Institutes of Health and is available through our *ANTsR* package. In subsequent sections we describe in greater detail the automated supervised brain segmentation pipeline including the generation of asymmetry feature images and the concatenated RF construction. We also describe the code and data used for the BRATS challenge so that the interested reader can reproduce our results. Finally, we illustrate performance with results from the BRATS 2013 challenge.

Materials and Methods

Feature generation for the proposed framework is critical. The performance of the selected machine learning strategy is limited by the input used for training. The image features described below were selected based on previous work in the literature and our own observations regarding tumor characteristics and what could possibly be encoded algorithmically. Observations regarding tumor intensity differences from normal tissue has been encoded in previously proposed research by calculating normalized neighborhood intensity statistics and intensity modeling (e.g., Bauer et al. 2011, 2012; Geremia et al. 2012; Zikic et al. 2012). In these approaches both the features and classification algorithm are limited to voxelwise considerations.

Our success in the MICCAI 2013 challenge was largely due to the inclusion of an extremely important set of features based on the common observation of bilateral asymmetrization caused by tumor presence. Additionally, several features were designed to encode spatial coherence even though the RF classification is performed on a voxelwise basis. For example, connected component geometry of the intensity modeling labels were used to discriminate obser-

vations of compact, isotropic objects (vs. highly anisotropic) typical of tumor presence and growth (Greenspan 1972). Table 1 provides a listing of image features used and the motivation for their inclusion.

Certain key elements characterize the proposed supervised brain tumor segmentation protocol including:

- construction of symmetric multivariate templates,
- generation of image features including those based on:
 - asymmetry and template normalization,
 - intensity modeling and connected component geometry,
 - neighborhood first-order statistics, and

– brain mask coordinate systems.

- training and geometric refinement of RF models, and
- prediction using the proposed framework.

In subsequent sections, we describe these items and detail the ANTsR framework which coordinates all steps.

Optimal Symmetric Multivariate Templates

In order to better characterize deviations from normal brain shape and appearance, several image features were derived using symmetric population-specific multivariate templates. For normal neuroanatomy, the use of spatial prior

Table 1 Image feature list for the proposed supervised brain tumor segmentation.

Asymmetry and template normalization		
Feature	Number	Motivation
contralateral difference	1 per modality	contralateral comparison
template difference	1 per modality	average normal comparison
log Jacobian	1	tumor distortion
warped template distance	1	tumor distortion
Intensity modeling and connected component geometry		
Feature	Number	Motivation
Pr(CSF)	1 per modality	normal tissue intensity
Pr(gray matter)	1 per modality	normal tissue intensity
Pr(white matter)	1 per modality	normal tissue intensity
Pr(edema)	1 per modality	normal tissue intensity
Pr(non-enhancing)	1 per modality	tumor tissue intensity
Pr(enhancing)	1 per modality	tumor tissue intensity
Pr(necrosis)	1 per modality	tumor tissue intensity
elongation	1 per modality	anisotropic components
eccentricity	1 per modality	anisotropic components
volume	1 per modality	small, isolated components
distance to tumor core	1 per modality	proximity to tumor core
$\frac{\text{volume}}{\text{surface area}}$	1 per modality	anisotropic components
Neighborhood first-order statistics		
Feature	Number	Motivation
mean (radius = 1)	1 per modality	tissue & tumor intensity
std. dev. (radius = 1)	1 per modality	tissue & tumor intensity
skewness (radius = 1)	1 per modality	tissue & tumor intensity
mean (radius = 3)	1 per modality	tissue & tumor intensity
std. dev. (radius = 3)	1 per modality	tissue & tumor intensity
skewness (radius = 3)	1 per modality	tissue & tumor intensity
T1, T1C difference	1	isolate tumor tissue
Brain mask coordinate system		
Feature	Number	Motivation
subject distance	1	peripheral vs. subcortical

information coupled with image normalization capabilities has proven useful in producing improved segmentation results of “expected” brain tissue such as cerebrospinal fluid, gray matter, and white matter (e.g., Ashburner and Friston, 1997). In contrast, accommodating spatial priors to model the presence of a possible tumor and its constituent tissue components is difficult. However, since the normal brain exhibits a bilaterally symmetric organization, we can use the presence of asymmetries to potentially differentiate abnormal brain tissue. A similar motivation prompted the identification of the mid-sagittal plane of symmetry (Prima et al. 2002) for feature generation in multiple sclerosis lesion (Geremia et al. 2011) and tumor (Geremia et al. 2012) identification. However, this earlier approach does not take into account the displacement of normal tissue due to tumor growth causing the mid-sagittal plane to deform from its planar structure (cf Fig. 1).

To take into account these potential asymmetries and the distortion of the normal plane of symmetry in the brain, we require a data set with the same modalities as dictated by the subject image acquisition protocol. Although it is preferable to build population-specific multivariate templates from normal data using the same acquisition parameters (Avants et al. 2010), such data were not available from the BRATS challenge. Therefore, we substituted well-known, publicly available data from a recent neuroimaging reproducibility study by Landman et al. which resulted in an open data cohort of 21 normal individuals, each imaged twice, comprising several modalities including ASL, FLAIR, DTI, fMRI, T1, and T2 (Landman et al. 2011). These data (known as the “MMRR” data set) were selected for deriving a multivariate template due to its public availability and inclusion of several modalities even permitting future incorporation of modalities not currently included with the BRATS challenge into our segmentation framework.

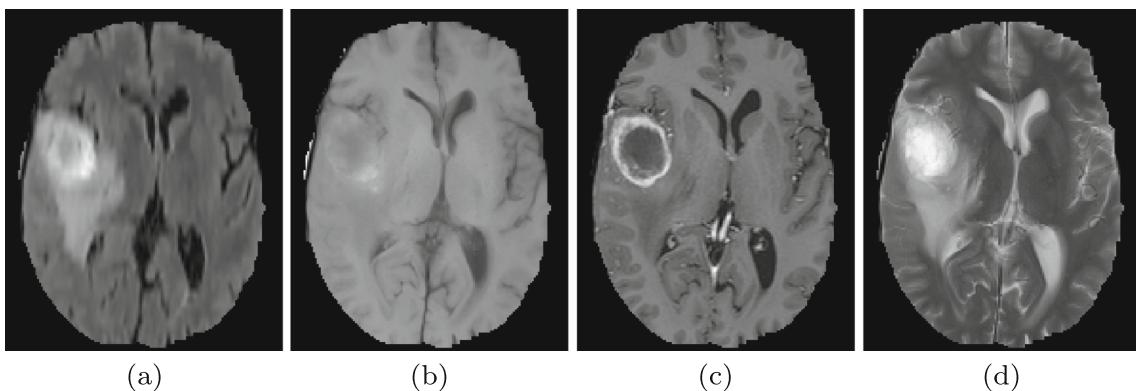


Fig. 1 Induced bilateral asymmetry due to tumor presence causing distortion of the plane of symmetry. Shown are mid-axial slices of one of the BRATS 2013 training data (specifically BRATS_HG0001): (a) FLAIR, (b) T1, (c) T1C, and (d) T2

As detailed in Avants et al. (2008a, 2010) given K multimodality images, $\mathbf{I} = \{I_1, I_2, \dots, I_K\}$, for N subjects, multivariate template construction iterates between optimizing the set of diffeomorphic transforms between the subjects and the template, $\{\(\phi_1, \phi_1^{-1})\}, \dots, (\phi_N, \phi_N^{-1})\}$ and constructing the optimal multivariate template appearance $\mathbf{J} = \{J_1, J_2, \dots, J_K\}$, with corresponding coordinate system $\psi(\mathbf{x})$, to minimize the following cost function:

$$\sum_{n=1}^N \left[D(\psi(\mathbf{x}), \phi_1^n(\mathbf{x}, 1)) + \sum_{k=1}^K \lambda_k \Pi_k(I_k^n, J_k(\phi_n^{-1}(\mathbf{x}, 1))) \right]. \quad (1)$$

D is the diffeomorphic shape distance,

$$D(\phi(\mathbf{x}, 0), \phi(\mathbf{x}, 1)) = \int_0^1 \|v(\mathbf{x}, t)\|_L dt \quad (2)$$

dependent on the choice of linear operator, L , and v is the velocity field

$$v(\phi(\mathbf{x}, t)) = \frac{d\phi(\mathbf{x}, t)}{dt}, \quad \phi(\mathbf{x}, 0) = \mathbf{x}. \quad (3)$$

Each pairwise registration employing the similarity metric Π_k can be assigned a relative weighting, λ_k , to weight a particular modality’s influence in the construction process. Once the multivariate template has converged (typically in four iterations), we symmetrize the template by flipping each asymmetric template component contralaterally and then running the multivariate template construction a second time using only the multivariate template and its symmetric analog. This is illustrated conceptually in Fig. 2.

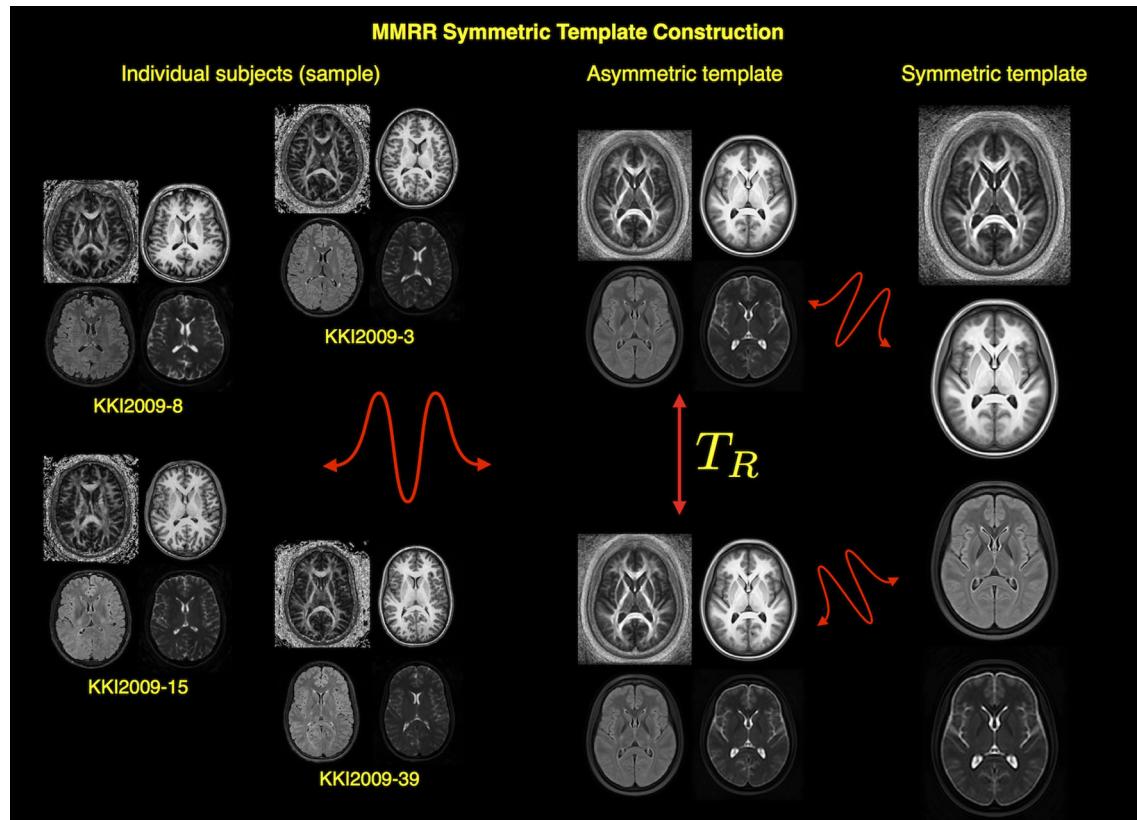


Fig. 2 Multivariate symmetric template created from the MMRR data (Landman et al. 2011). Of the seven modalities comprising the set of study acquisitions, we illustrate the (a) FA, (b) FLAIR, (c) MPRAGE, and (d) T2 template components. Although DWI-based images were not included in the challenge data, such images have shown discriminative potential (Price et al. 2003; Cha 2005) warranting investigation in our future work. The optimal transformation and averaging of the

individual subject images result in the asymmetric template represented at the top of the middle column. A horizontal reflection, T_R , perpendicular to the mid-sagittal plane resulted in the contralateral counterpart represented at the bottom of the second column. The final template seen on the right is a result of repeating the template construction using the two asymmetric templates as input

Image Features Based on Asymmetry and Symmetric Template Normalization

After constructing the template offline, each data set is processed by first registering the non-contrast T1-weighted image to the T1-weighted component of the symmetric template. To do this we use a recently developed SyN (Avants et al. 2011a) variant based on B-spline regularization which has demonstrated good performance in normal brain registration (Tustison and Avants 2013). We denote the mapping from the subject, S , to the template, T , space as $\phi_b : S \xrightarrow{b} T$ which consists of both affine and diffeomorphic components. Note that transform invertibility is essential for the template-based features. The first set of feature images is generated by warping the template components to the subject space and calculating the difference image (see the top portion of Fig. 3). Note that the T1-weighted contrast and non-contrast images are paired with

the T1-weighted component of the symmetric template. For example, the T2 symmetric template voxelwise difference image is calculated from

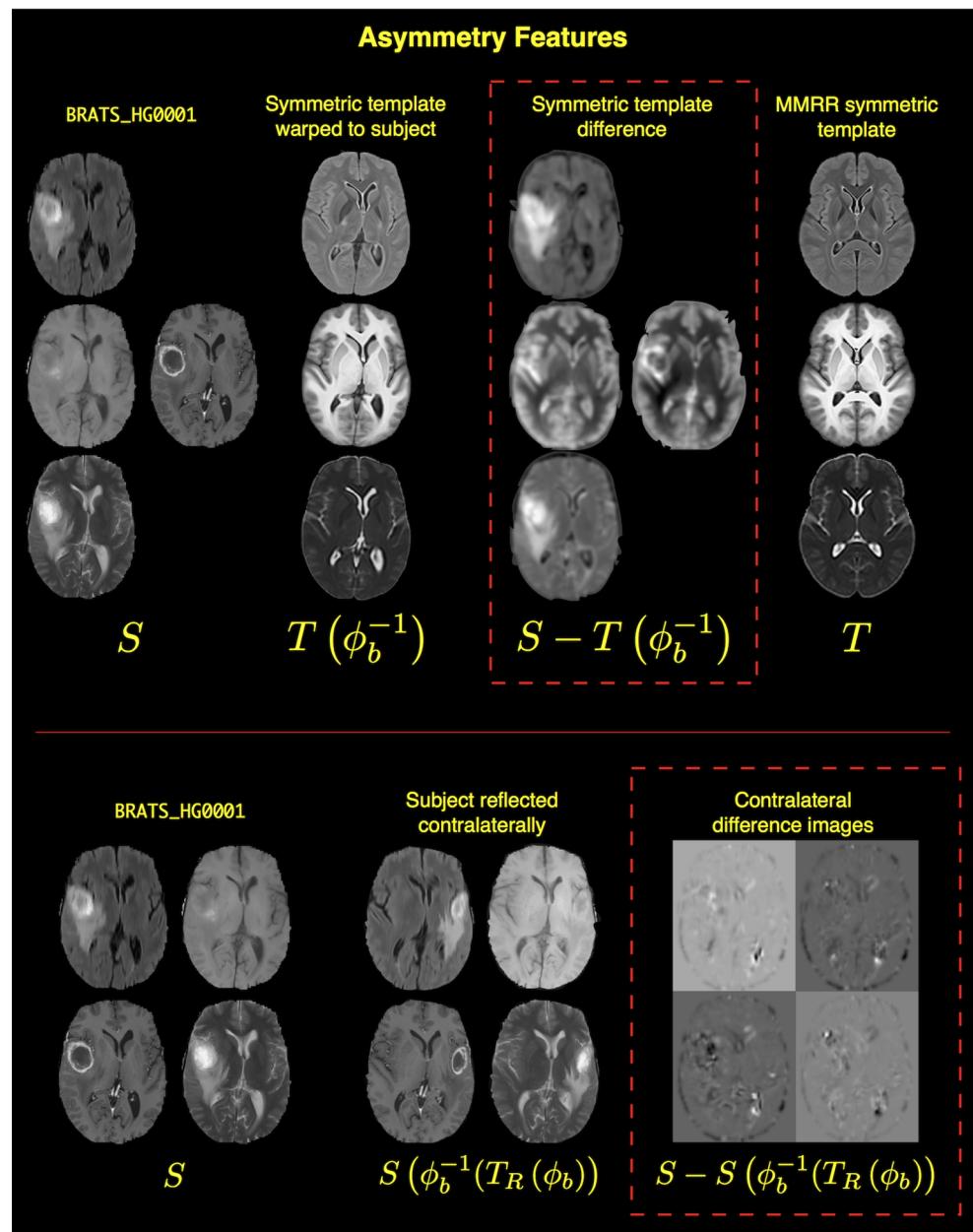
$$\text{T2 symmetric template difference} = S_{\text{T2}} - T_{\text{T2}}(\phi_b^{-1}). \quad (4)$$

The second set of template-based feature images is generated per modality as the difference image with the contralateral reflection. This is achieved by calculating the reflection transform, T_R , in the symmetric template space and composing transforms as follows to create the non-Euclidean contralateral counterpart per modality:

$$S_{\text{contralateral}} = S(\phi_b^{-1}(T_R(\phi_b))). \quad (5)$$

The corresponding feature image is calculated as the voxelwise difference $S_{\text{contralateral}} - S$. This process is illustrated in the bottom portion of Fig. 3. Two related features are the Jacobian determinant image of the transform (calculated voxelwise) from the template to the subject, i.e., $J(\phi_b^{-1})$ and

Fig. 3 Given the mapping between the template, T , and subject, S , domains ($\phi_b : S \xrightarrow{b} T$), various features can be calculated which demonstrate good discriminative qualities. T_R denotes a horizontal reflection perpendicular to the mid-sagittal plane. Feature images used are specified by the dashed box. Top: Difference images with the symmetric multivariate template are created by warping the template to the subject space and performing a voxelwise subtraction from the original modality image. Bottom: Similarly, contralateral difference images are calculated from each modality per subject by generating the non-Euclidean contralateral image via the diffeomorphic transform ϕ_b



the warped signed distance template mask (Maurer et al. 2003). The motivation for inclusion of these two feature is that relatively larger Jacobian values are a potential indicator of tumor expansion similar to the distorted distance map (Fig. 4).

Voxelwise Image Features for Random Forest Supervised Segmentation

We use random forests as supervised learners. Given ground truth, voxel-wise n -class labels for a specific image, we can vectorize the label image to produce v_L , the p -length vector

of class labels where p is the number of voxels in the given brain of interest. We also assume a matrix of features for the given brain, which we denote F where F has dimensionality of p by k , where each of the k columns corresponds to a potentially valuable feature vector. Given this representation, the random forest model can be trained, in R notation, to predict the labels via $v_L \sim F$. Our segmentation protocol involves training and application of two RF models in succession (see Fig. 5) to which we refer as “stages.” The basic idea is that we generate a set of feature images used as input to the first RF model (or first stage) which produces a voxelwise probabilistic tissue estimate. More precisely, in

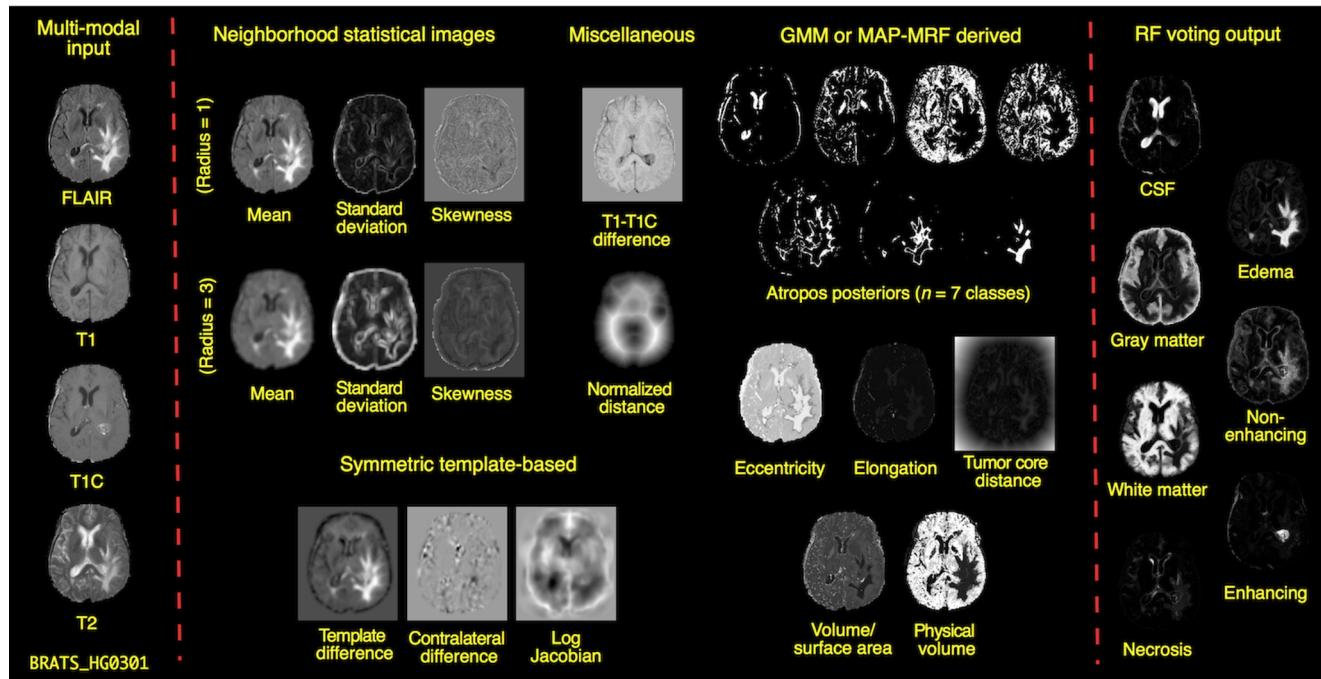


Fig. 4 Representative feature images derived from the BRATS_HG0301 final assessment data set. Neighborhood statistical images for each modality were generated by calculating a given statistic within a specified neighborhood radius. Similarly, we calculate the voxelwise difference image between the T1 and T1C image. Also calculated for each modality were feature images based on either the GMM or the MAP-MRF segmentation. For the former, we show the probability maps for each of the seven labels which are used as feature images. From the resulting hard segmentation, we calculate

various geometric measures per connected component of each of the seven labels. Similarly, the registration to the symmetric template produces the modality-specific difference images with the corresponding symmetric template itself and with respect to the contralateral side. This mapping is also used to produce the log Jacobian image and warped template distance mask. Additionally, the ($T1 - T1C$) image is calculated and, from the cerebral mask, we calculate the normalized distance image

the RF framework, each prediction sample (i.e. the feature vector at each voxel), is propagated through each tree of the ensemble where it is labeled as belonging to a specific class. These “votes” are converted to voxelwise probabilistic estimates for each class via standard mechanisms used in random forest models, e.g. Liaw and Wiener (2002). We then use these output tissue map estimates as spatial priors for generating a second set of geometrically refined image features. These are used as input for a second RF model application (or second stage), the output of which constitute the final tumor segmentation estimate. Note that some feature images are included directly and without modification in both stages such as the asymmetry features described in the previous section.

Image Features Based on Intensity Modeling and Regional Geometry

Intensity modeling of the constituent tissue types has proven useful in previous tumor segmentation protocols. For example, Gaussian mixture modeling (GMM) was used in Bauer

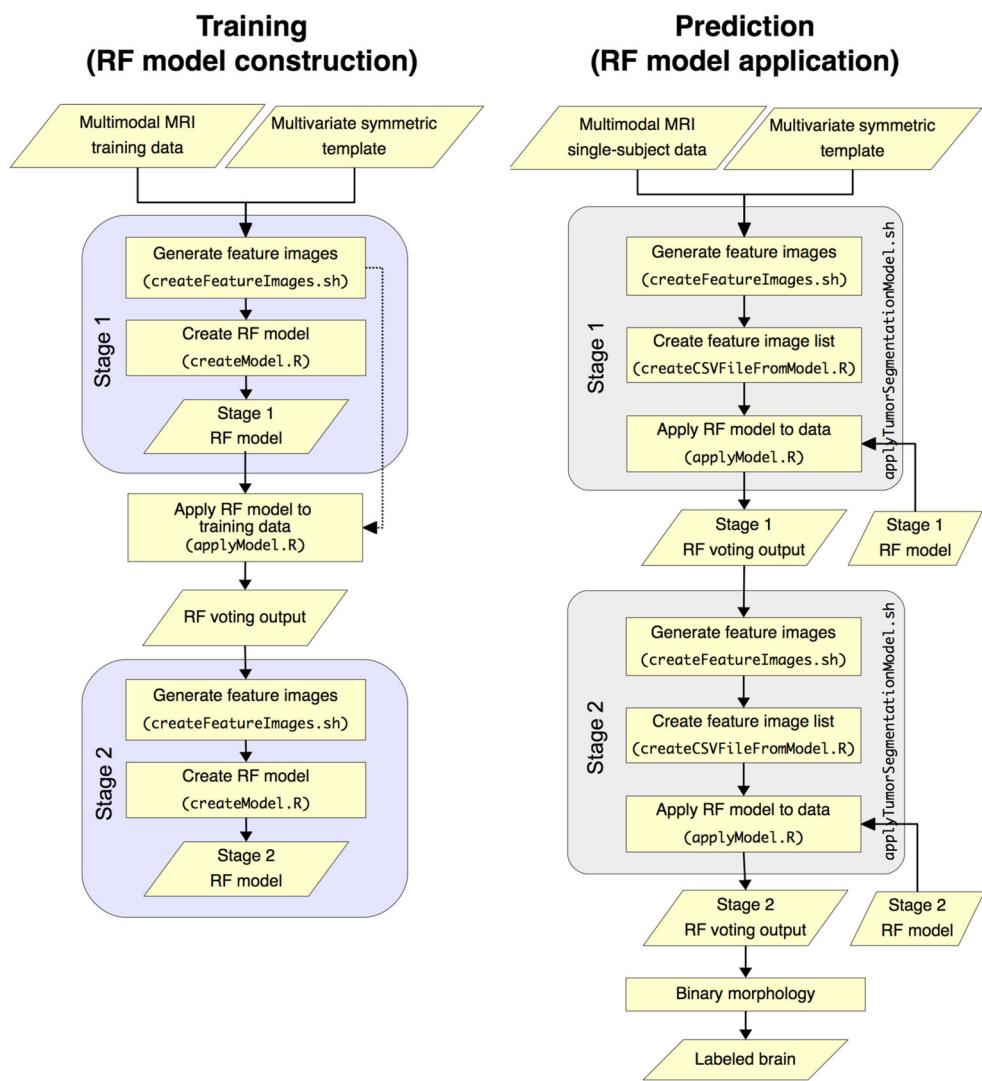
et al. (2012) and Zikic et al. (2012) to model tumor tissue components. Similarly, we use GMM to model the seven brain/tumor tissue types:

- cerebrospinal fluid,
- grey matter,
- white matter,
- edema,
- non-enhancing tumor (including low-grade tumor center),
- enhancing tumor (excluding necrotic center), and
- abnormal necrotic center or necrocyst in high-grade gliomas.

However, since GMM is performed without any spatial considerations, we augment the initial supervised segmentation classification result by employing a second classification round using an MAP-MRF prior on the same seven brain/tumor tissue types (detailed below).

In contrast to previous generative modeling approaches for multimodal tumor segmentation (e.g., Prastawa

Fig. 5 Diagrammatic workflow for the proposed RF model training (left) and prediction (right). For training, the feature images are first generated from the set of input training multimodal MRI and symmetric multivariate template. The set of features images are used to create the RF model for the first stage. This first stage RF model is then applied to the set of input training data to yield spatial priors for seeding the generation of the second stage feature images. These feature images are then used to create the second stage RF model. For prediction of a single-subject segmentation, a similar scenario is applied whereby the generated feature images are used as input to the RF models the second of which produces an initial segmentation estimate. This estimate is then refined using a series of heuristically-derived binary morphological operations resulting in the final labeled brain



et al. 2003), we do not use multivariate Gaussians to specify tissue probabilities but rather incorporate each univariate probability map into the feature vector of the training data. As pointed out in Menze et al. (2010), parametric multivariate modeling might obscure the distinct biological information provided by each modality. Instead, we let the RF construction process determine the optimal combination of such multivariate information. Additionally, maximum posterior labeling from both stages is used to determine the connected components for each label. Geometric features (assigned voxelwise) include the physical volumes of each connected component, the volume to surface area ratio, the elongation, and eccentricity. Note that random forests are known to be robust to the presence of multiple correlated features.

Stage 1: Voxelwise Classification with Gaussian Mixture Modeling The first step in processing a new image is to

segment the image into the n tissue classes listed above. We achieve this via an expectation-maximization segmentation method based on GMM. GMM results in voxelwise tissue probabilities which comprise an additional set of image features. The GMM is initialized with prior cluster centers for specific tissue types. We learn the values for these cluster centers for each modality and each tissue from training data (Reynolds 2009). We also perform intensity normalization across the training cohort by winsorizing the intensity values to the quantile range [0.01, 0.99] and then rescaling the resulting intensities to [0, 1]. The cluster centers are defined as the mean normalized intensity value for each tissue type of each modality image over all the training data. We tried a more sophisticated intensity normalization scheme using the approach of Nyúl et al. (2000) but found that our simpler normalization approach, in combination with the rest of the pipeline, produced slightly better results.

More formally, the GMM computes the probability distribution at each voxel, \mathbf{x} , as the sum of M Gaussian components, $\mathcal{N}(\mathbf{x}|\mu, \sigma)$, i.e.

$$p(\mathbf{x}|\mu_m, \sigma_m, \lambda_m) = \sum_{i=1}^M \lambda_m \mathcal{N}(\mathbf{x}|\mu_m, \sigma_m) \quad (6)$$

where $\sum_{m=1}^M \lambda_m = 1$. The parameters of the GMM are determined using the Atropos segmentation tool (Avants et al. 2011b) available in ANTs. For the BRATS data, GMM modeling was applied to each of the four modalities, i.e., FLAIR, T1, T1 contrast, and T2 (see Fig. 6), separately. Note that no spatial priors are used during this stage including MRF spatial priors (Fig. 7).

The stage 1 random forest training uses the asymmetry features, the seven posterior probability tissue images output by the GMM along with five additional geometric features based on the connected components of the segmentation image. These are created by first isolating each of the seven tissue labels. For each label we determine the connected components then, for each connected component, we calculate its physical volume and volume-to-surface ratio. Additionally, we calculate the component's second-order central image moments assembled into the image covariance matrix (Padfield and Miller 2008). The ordered eigenvalues are denoted as $\{\lambda_1, \lambda_2, \lambda_3\}$ and are used to calculate a 3-D estimate of the eccentricity,

$$\text{eccentricity} = \frac{\lambda_3 - \lambda_1}{\lambda_3}, \quad (7)$$

and elongation,

$$\text{elongation} = \frac{\lambda_3}{\lambda_1}, \quad (8)$$

meant to discriminate isotropic vs. anisotropic objects. Note that each measure applies to a single connected component but the value is assigned to each voxel comprising that labeled object. For each image, we also calculate the Euclidean distance to the tumor core (label 7).

During the training phase (or RF model construction), these feature images, $\{F_i : i \in GMM\}$, generated during Stage 1 are used to create the Stage 1 RF model with the following relationship

$$v_L \sim \sum_{i \in Asym} F_i + \sum_{i \in GMM} F_i + \sum_{i \in Neigh} F_i + \sum_{i \in Dist} F_i \quad (9)$$

given in the standard R notation of Wilkinson and Rogers (1973). In other words, $\mathbf{F} = [\mathbf{F}_{Asym}, \mathbf{F}_{GMM}, \mathbf{F}_{Neigh}, \mathbf{F}_{Dist}]$, where we column-wise concatenate the asymmetry ($Asym$), intensity modeling and connected component geometry (GMM), neighborhood

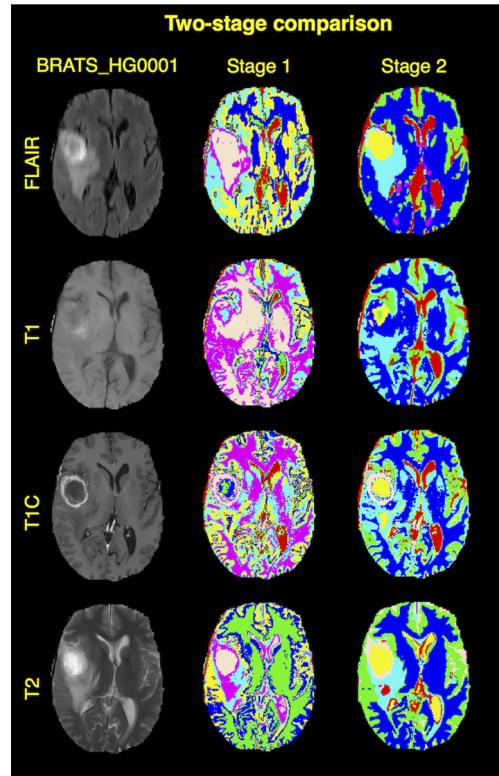


Fig. 6 Visual comparison between the first and second RF stages on the evaluation BRATS_HG0001 data set. Although the GMM intensity modeling does provide certain discriminative benefit (see Figs. 8 and 9), the use of the RF-derived posteriors coupled with the described MAP-MRF segmentation framework enhances the accuracy of the intensity-based modeling features

statistics ($Neigh$), and brain mask signed distances ($Dist$) feature matrices.

Application of the Stage 1 RF model to a set of feature images from an individual subject (e.g., Fig. 4) results in a set of n spatial images containing the voxel-wise probabilities for each of the seven tissue types. At each voxel, the corresponding multi-feature vector is propagated through each tree of the random forest resulting in a single vote/classification per tree. The classification probabilities are derived by normalizing these votes to [0, 1] (Liaw and Wiener 2002).

Stage 2: Refined Spatial Coherence Classification Initialization of stage 2 employs the per-class probabilities output by the stage 1 RF model as input to an expectation-maximization MAP-MRF segmentation algorithm. Specifically, these spatial probability maps are passed as priors to the `antsAtroposN4.sh` script in ANTs which couples the Atropos segmentation tool (Avants et al. 2011b) with N4 bias correction (Tustison et al. 2010). The posterior probabilities that are output by Atropos and the connected-component feature images for each modality are denoted as

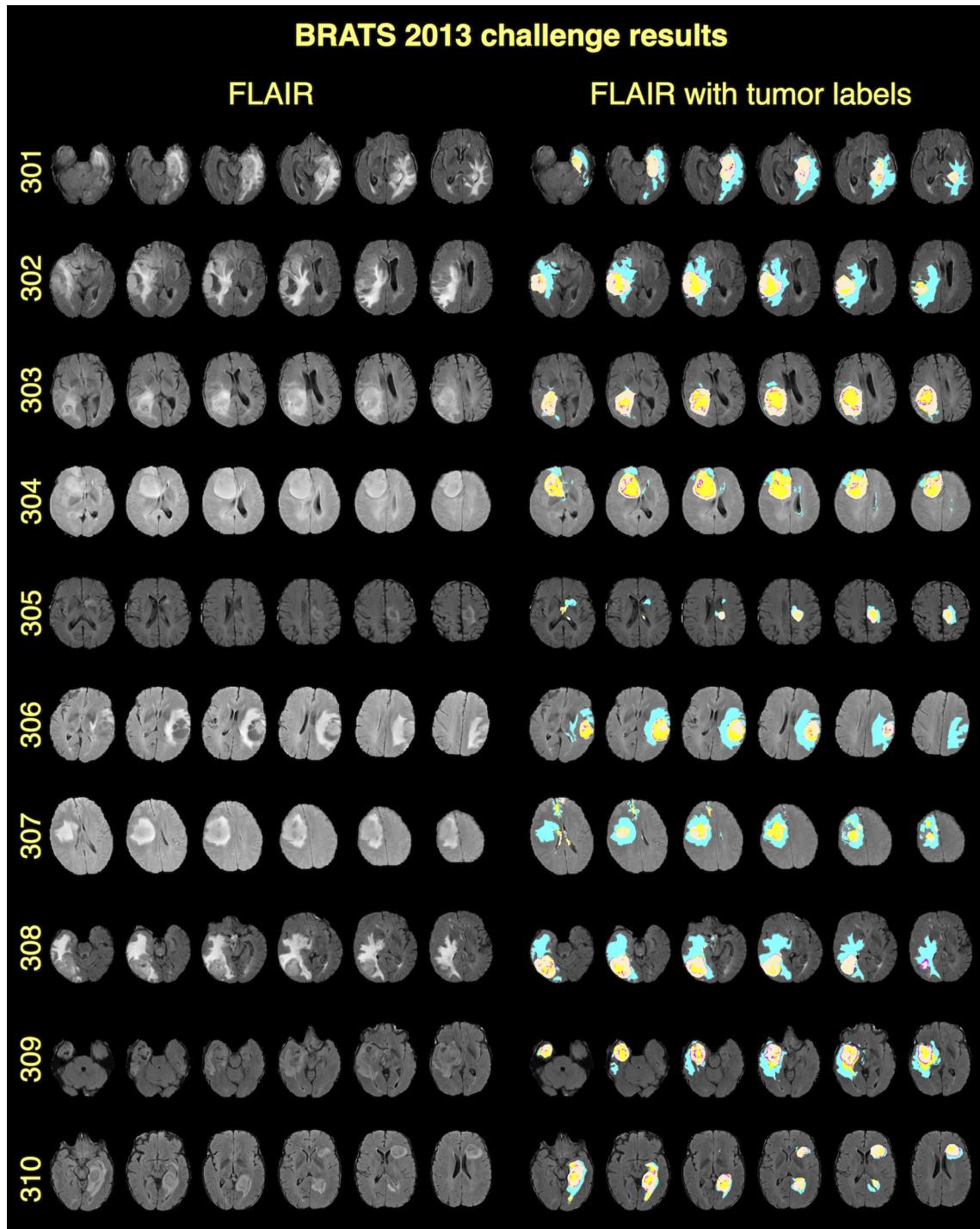


Fig. 7 Visual results (every tenth slice sampled from the axial center of the tumor region) from the BRATS 2013 challenge using the proposed framework (off-white = non-enhancing tumor, cyan = edema, violet = enhancing tumor, yellow = necrosis). Tumor components were

combined into three regions for performance assessment ((1) complete tumor: all four tissue classes; (2) tumor core: necrosis, non-enhancing tumor, and enhancing tumor; and (3) enhancing tumor). See Table 2 for quantitative measures and challenge rankings

\mathbf{F}_{MRF} . The second stage RF model is then produced similarly as the previous stage using the concatenated feature set $\mathbf{F} = [\mathbf{F}_{Asym}, \mathbf{F}_{MRF}, \mathbf{F}_{Neigh}, \mathbf{F}_{Dist}]$

$$v_L \sim \sum_{i \in Asym} F_i + \sum_{i \in MRF} F_i + \sum_{i \in Neigh} F_i + \sum_{i \in Dist} F_i. \quad (10)$$

The use of these MRF features improves the accuracy of the intensity-based modeling over the GMM approach by introducing a spatial coherence constraint (the MRF prior) to those features in addition to providing a better initialization with the RF posteriors from the initial estimation stage

(which lacked the MRF spatial prior). In total, the number of features is for this final stage is the same where the GMM features are replaced by their MRF counterparts. The rest of the feature images from the first stage are also included unaltered for this second stage.

ANTsR: An ANTs/R Interface for Random Forest Training and Prediction

The complexity of neuroimaging research necessitates commensurable numerical analysis capabilities. Similarly, concomitant with the era of “big data” (specifically with respect to neuroimaging (VanHorn and Toga 2013)) are new visualization needs and challenges (Childs et al. 2013; Kehler and Hauser 2013). In response, various software packages have been developed to integrate tools specific to neuroimaging research with more general numerical and visualization software packages. The well-known neuroimaging package SPM is a significant extension of the commercial computing and visualization environment Matlab. Open source neuroimaging packages, such as NIPY (neuroimaging in Python), rely on other open source packages for numerical/statistical analysis. NIPY, for example, uses the more generic packages NumPy and SciPy for numerical analysis and optimization.

ANTs (Advanced Normalization Tools) was built, originally, to provide high performance image registration for medical image analysis (Avants et al. 2008b) and based upon the mature Insight Toolkit (ITK) sponsored by the National Institutes of Health. Since then, ANTs has grown to include several robust medical image analysis solutions including bias correction (Tustison et al. 2010), *n*-tissue multivariate segmentation (Avants et al. 2011b), template construction (Avants et al. 2010), and cortical thickness estimation (Das et al. 2009) (many of which have been introduced into ITK partially in an attempted leveraging of Linus’s Law—“Given enough eyeballs, all bugs are shallow”). However, in the evolution of the toolkit, it became clear that robust statistical machinery was lacking for making inferences regarding the data produced during the course of ANTs processing. *ANTsR* was developed specifically to provide an interface between ANTs, a powerful neuroimaging toolkit for producing reliable imaging data transformations, and the *R* project for statistical computing and visualization thus providing a complete set of tools for multivariate neuroimage analysis. *ANTsR* intends to provide a modern framework for medical analytics, with a focus on imaging-assisted prediction and statistical power.

Careful consideration of available statistical software led to the adoption of *R* to complement ANTs quantification resulting in the *ANTsR* package. *R*’s open source code base, reliable software testing and distribution strategies, and

add-on packages coupled with its rapidly growing community of developers and users has caused wide-scale adoption within both academia and industry.

Installation

The *ANTsR* package is publicly available on the github project hosting service.¹ Prior to installation of *ANTsR*, several external *R* packages need to be installed including: *Rcpp*, *signal*, *timeSeries*, *mFilter*, *doParallel*, *robust*, *magic*, *knitr*, *pixmap*, *rgl*, and *misc3d*. Additionally, in order to perform the supervised brain segmentation as described in later sections, one needs to also install the packages *randomForest*, *snowfall*, *rlecuyer*, and *ggplot2*.

In addition to *R* and the add-on packages previously mentioned, CMake is also required. CMake is an open source tool for the management and building of large-scale software projects. It is used to coordinate the downloading of external packages, such as the Insight Toolkit (ITK) and ANTs. Further instructions for download and installation can be found on the *ANTsR* github website. Feel free to contact the authors if installation trouble occurs. We note that *ANTsR* is currently only tested on OSX and certain Linux operating systems.

Usage

ANTsR is intended to not only allow easy interchange between medical imaging formats and *R* but also to facilitate reproducible scientific studies and the type compilable analysis articles that are fundamental to journals such as *Biostatistics*. Both *knitr* and *sweave* facilitate integration of R-code with the LaTeX document preparation system.

An additional motivation for our development of *ANTsR* (and hopefully its acceptance by the community) stems from the ability to couple ANTs core functionality, including IO tools such as *antsImageRead*, with the large number of *R* statistical and visualization packages. Due to this combination, several functions have been easily created for such neuroimaging-specific tasks as fMRI/ASL data manipulation and analysis, voxel and ROI-based analyses, and connectivity visualization. The user help menu and documentation for the library and its constituent functions are invoked in the similar manner as other R libraries.

As mentioned earlier, we have made this entire framework available as open source. In addition to the *ANTsR* repository already on github which houses both ANTs and *ANTsR* functionality, we created a special github repository specifically for this work containing figures, references, and

¹<https://github.com/stnava/ANTsR>

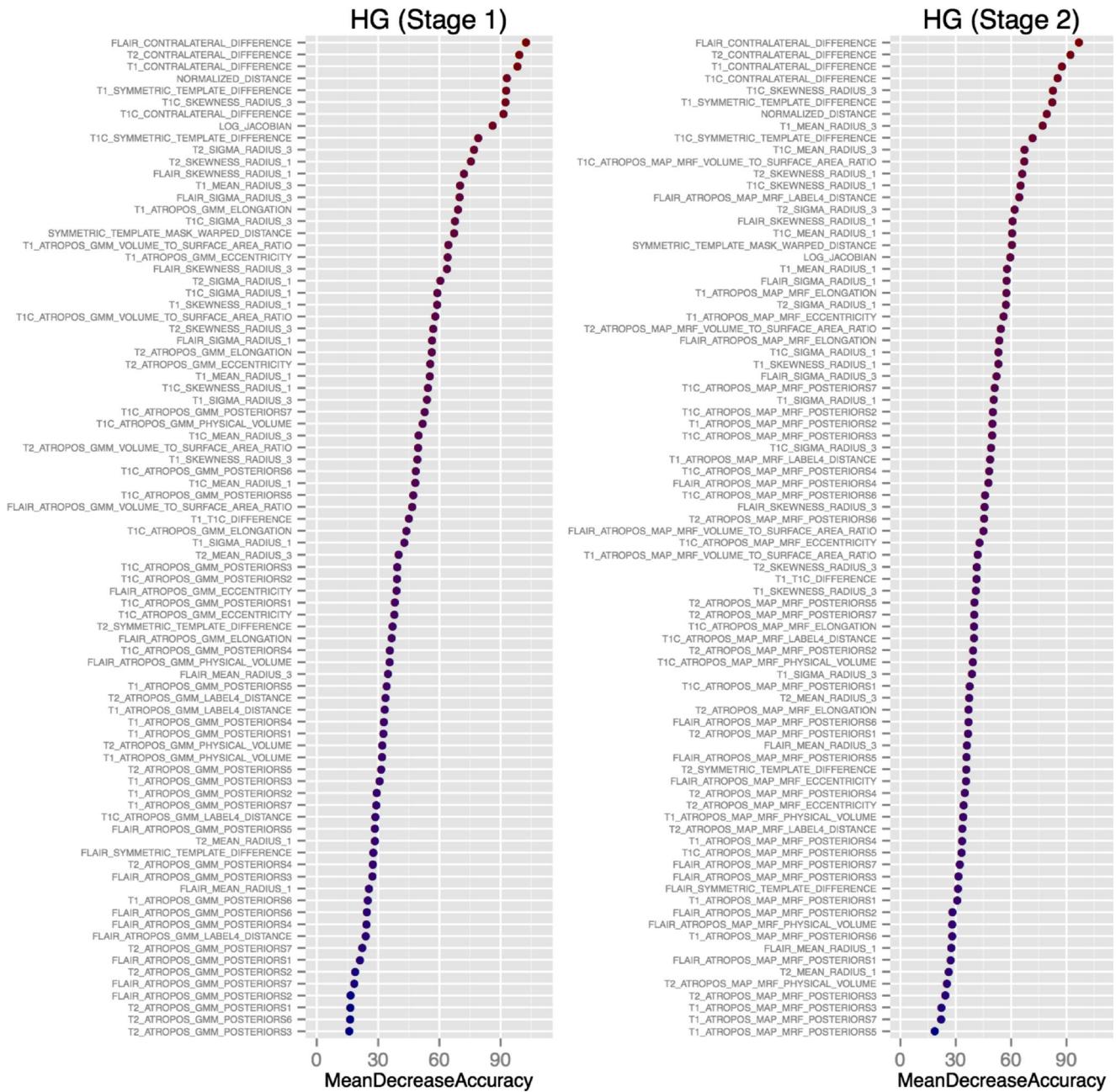


Fig. 8 MeanDecreaseAccuracy plots generated from the high-grade glioma Stage 1 and Stage 2 RF models. These plots provide a weighted ranking describing the importance of each feature for predictive accuracy

text.² Also, we posted all scripts (R, shell, and perl) used to coordinate the ANTsR processing including:

- applyModel.R: applies a RF model to a new feature data set from a testing subject resulting in a set of probability images (one for each label).

- applyTumorSegmentationModel.sh: generates the new feature image set from the testing MRI (by calling createFeatureImages.sh), organizes the file names in a csv file (via createCSVFileFromModel.R), and applies the RF model using applyModel.R.
- applyTumorSegmentationModelForCohort.pl: Coordinates tumor segmentation on the computational cluster for a given cohort.

²<https://github.com/ntustison/ANTsAndArboles>

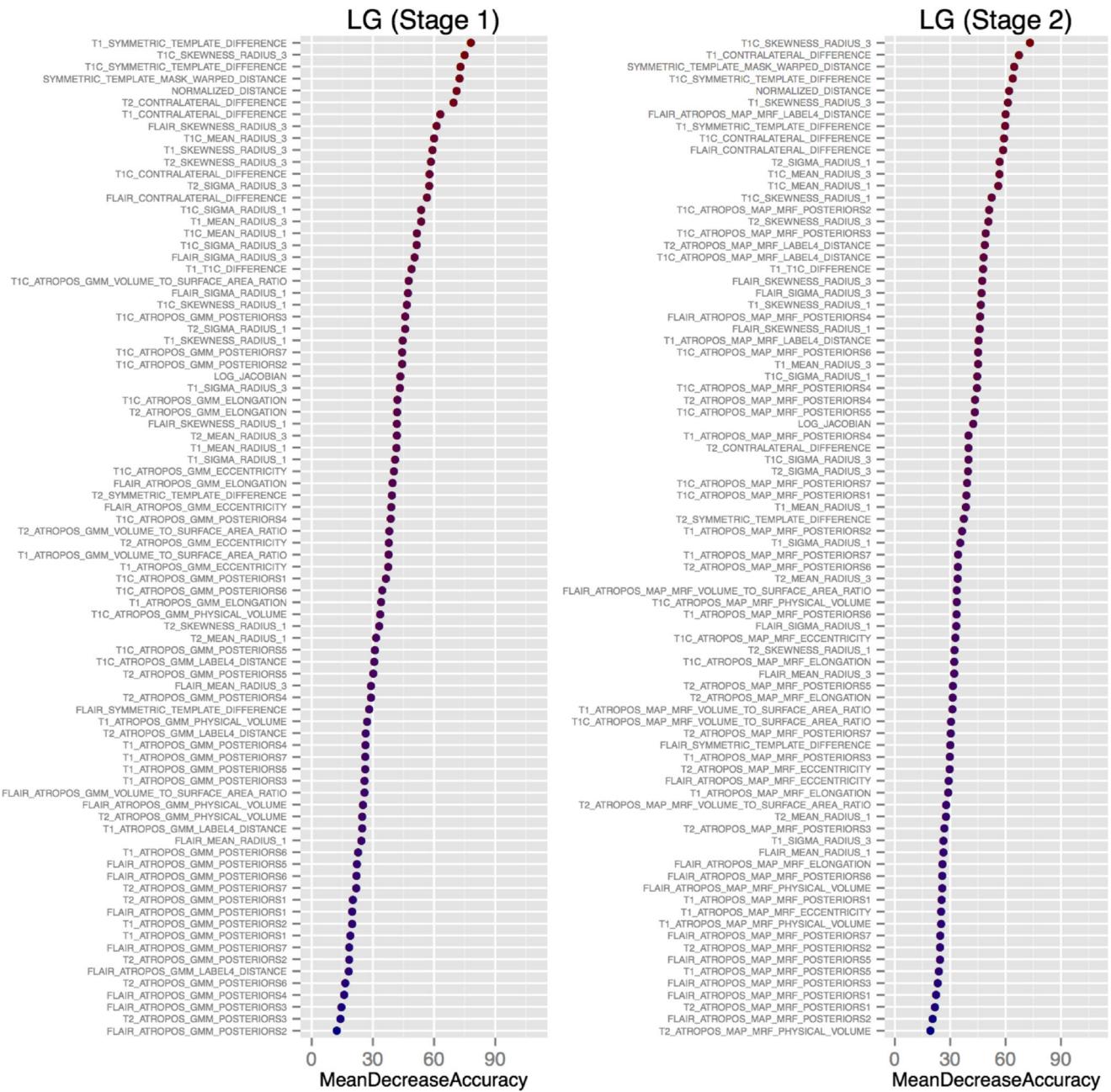


Fig. 9 MeanDecreaseAccuracy plots generated from the low-grade glioma Stage 1 and Stage 2 RF models. These plots provide a weighted ranking describing the importance of each feature for predictive accuracy

- `createCSVFileFromModel.R`: organizes the set of feature image file names in a csv file for input into `applyModel.R`.
- `createFeatureImages.sh`: creates the set of feature images given a set of co-registered input MRI from a single subject.
- `createModel.R`: creates a RF model given the input csv file of the feature image file names for all training data and set of truth label maps.

- `plotVariableImportance.R`: produces a plot of the importance of each feature variable used in constructing the model.

We also include a fully functional 2-D example which performs both testing and training on sample challenge data. After pulling the repository, one can run the scripts `exampleTrain.sh` and `examplePredict.sh` to get the sample results. Output includes several overlap measures describing the performance.

Brain Tumor Data

Brain tumor image data used in this work were obtained from the NCI-MICCAI 2013 Challenge on Multimodal Brain Tumor Segmentation organized by K. Farahani, M. Reyes, B. Menze, E. Gerstner, J. Kirby and J. Kalpathy-Cramer. The challenge database contains fully anonymized images from the following institutions: ETH Zurich, University of Bern, University of Debrecen, and University of Utah and publicly available images from the Cancer Imaging Archive (TCIA). Both training and testing data were made freely available through the Creative Commons Attribution-NonCommercial 3.0 license.

Training data consisted of multimodal brain MRI (T1, T2, FLAIR, and post-Gadolinium T1) from 30 glioma patients (both low, $n = 10$, and high-grade, $n = 20$, and with and without resection). For each subject, the T1, T2, and FLAIR MRI were linearly registered to the post-contrast T1. Subsequently, the brains were skull-stripped and resampled to 1 mm isotropic resolution. Testing data was processed similarly and released during the course of the challenge in two sets denoted as “Leaderboard” and “Challenge” data. The former consisted of 21 and 4 high and low-grade tumor patients, respectively, whereas the latter comprised 10 high-grade only patients.

Manual labeling was performed in the axial plane following a detailed protocol. The labeling of pathology was categorized into four regions: edema, non-enhancing tumor (including low-grade tumor center), enhancing tumor (excluding necrotic center), and abnormal necrotic center or necrocyst in high-grade gliomas. Normal brain tissue was not labeled.

Training: RF Creation for the BRATS 2013 Challenge

] For use with the Challenge and Leaderboard data, cohort-specific models (both low-grade and high-grade glioma) for both GMM and MAP-MRF stages were created using only the supplied training data. Prior to training, we segmented normal brain tissue (Avants et al. 2011b) for each data set by segmenting only the T1 image. This was only to yield a rough estimate of normal brain tissue to augment the already provided pathology labels. This resulted in seven labels for tissues described earlier i.e., csf, gray matter, white matter, necrosis, edema, enhancing, and non-enhancing tumor characterizing each brain.

Initial testing of our proposed framework was performed on the training data using a leave-one-out strategy. Once the feature images are created for each subject, the resulting images of the entire training cohort are organized in a csv file for input into the R script `createModel.R`.

Other possible input parameters include the requested number of trees, number of samples per label, and number of threads for parallel processing. The output is an RData file describing the RF model which can be used for future predictions.

Additionally, the `randomForest` package provides measurements for determining the importance of chosen features when constructing the model. This aids in potential feature pruning or intuiting model behavior. As mentioned earlier, we provide the R script `plotVariableImportance.R` to render one such quantity use to assess RF importance using `MeanDecreaseAccuracy`. During model construction (specifically the out-of-bag error calculation stage), the decrease in prediction accuracy with the omission of a single feature or variable is tracked and averaged. Thus, those features which have the greatest decrease in mean accuracy are considered to be the most discriminative. In this work, we do not use these measurements for feature pruning. However, we plot them in the Results section (see Figs. 8 and 9) as they demonstrate the relative importance of our selected features including that of the proposed asymmetry images.

Prediction: Applying the RFs for the BRATS 2013 Challenge

Once the models are created, classification of tumors in new subjects is performed as illustrated in Fig. 5. From the feature images and input GMM model, a tentative set of RF voting output confidence images are produced. As described, this is used as input to the second prediction round. The final probability output images are used to produce the maximum probability labeling.

A final round of binary morphological operations were heuristically designed to improve the final segmentation results such as removal of small connected components and morphological closure of certain regions. All steps are included in the script `applyTumorSegmentationModelForCohort.pl` designed for parallel subject processing on the computational cluster at the University of Virginia.

Results

A total of four RF models were created from the 30 training data sets. Two models for Stage 1 and Stage 2 processing were generated from the 20 high-grade glioma evaluation data described earlier. Similarly, two additional models were created from the 10 low-grade glioma data sets.

Visual results from the challenge data set consisting exclusively of high-grade gliomas are given in Fig. 7. Following model construction, weighted importance feature plots described earlier were rendered to provide feedback as to the individual potential predictive accuracy of each feature. The high-grade plots are given in Fig. 8 whereas the low-grade plots are given in Fig. 9.

Immediately apparent from these plots are the importance of certain features. In general, the features based on the symmetric template are quite discriminative thus justifying the increased computational resources required to generate such features. For single-threaded processing on the cluster, creating the feature images for a single subject required approximately 2 hours of total processing time of which the template registration component took approximately 75%. Additionally, the first order statistical features also seemed fairly important which is something that previous work had demonstrated (e.g., Bauer et al. 2012; Zikic et al. 2012). There are a number of differences, however, between low-grade and high-grade importance outcomes which would also seem to justify the creation of separate glioma class models.

As described earlier, the MICCAI 2013 BRATS challenge data was provided to the competitors in three sets. The Evaluation set was used primarily for training although an initial ranking was performed for all the competitors to aid in determining participation in the workshop. Shortly prior to the actual workshop date, the Leaderboard data was released to the competitors for posting of results and subsequent ranking. Finally, the night before the competition, the Challenge data was made available and used to produce the final competitor ranking. In addition to the Dice overlap measure, additional performance measures for producing the rankings included the positive predictive value, and

sensitivity (all of which can be calculated using open source tools such as Tustison and Gee (2009)). For all three data sets, we provide these overlap measures and relative competitor ranking in Table 2. Full competition results can be viewed at <http://www.virtualskeleton.ch>.

Discussion and Conclusions

One of the difficulties with the focus of this particular competition is the lack of consensus, even clinically, as to tumor type and extent in the context of medical imaging (Cha 2005). In fact, current consensus guidelines from the World Health Organization for brain tumor classification are strictly histopathological (Louis et al. 2007) which limit clinical application. Such limitations motivate the use of medical imaging for treatment planning and outcome analyses (Cha 2005) including more automated methods such as that proposed in this work.

Although previous research has employed RFs for supervised brain segmentation, our contribution of concatenated RF model application and use of symmetric multivariate templates demonstrated good performance in the recent MICCAI 2013 Brain Tumor Segmentation challenge/workshop. In terms of variable importance, the latter provided several highly discriminative features which resulted in the top-performing algorithm of the competition. This confirmed what others have found in that RFs provide an excellent framework for prediction in certain medical image analysis problems. Even with relatively few training subjects relative to input variables, the RF models perform well. In addition, this work highlights the value of a symmetric multiple modality template for clinical disorders in which asymmetry is a hallmark. Furthermore, our

Table 2 Assessment measures from the MICCAI 2013 BRATS Challenge including Dice overlap, positive prediction value (PPV), and sensitivity.

Data Set	Rank [†]	Dice	PPV	Sensitivity
Challenge (complete)	1	0.87 (1)	0.85 (2)	0.89 (2)
Challenge (core)		0.78 (1)	0.74 (5)	0.88 (1)
Challenge (enhanced)		0.74 (1)	0.69 (4)	0.83 (1)
Leaderboard (complete)	1	0.79 (2)	0.83 (1)	0.81 (4)
Leaderboard (core)		0.65 (1)	0.70 (2)	0.73 (2)
Leaderboard (enhanced)		0.53 (3)	0.51 (3)	0.66 (2)
Evaluation (complete)	2	0.88 (2)	0.88 (3)	0.89 (3)
Evaluation (core)		0.76 (2)	0.80 (5)	0.79 (3)
Evaluation (enhanced)		0.55 (3)	0.65 (3)	0.53 (3)

[†] Indicates total performance over all regions. Component rankings are in parentheses

challenge-leading results establish the value of such templates in multiple modality prediction tasks in which there is a small training set with an abundance of multivariate data.

Objective assessment of algorithmic performance (given the prevalence of selection bias in reporting results in conventional publication venues) is certainly a motivating factor for these competitions. Equally, if not more, important is the availability of the actual algorithmic instantiation (i.e. code) so that others can more easily build upon and utilize what has proven effective. This also provides an opportunity for the users to return constructive feedback to the authors thereby improving the original offering.

However, as explained previously, the brain tumor segmentation methodology that we have made available is only a small part of the larger software package that we have created in *ANTsR*. Not only does *ANTsR* significantly facilitate the development of the work discussed, but it provides an interface to one of the most powerful statistical packages available in R. The combination of the well-known ANTs software package with R provides tremendous potential for future insightful analysis.

Information Sharing Statement

The open source ANTsR package (RRID:nlx_151365) is available as compilable source code or as a set of binaries from its dedicated github repository (<https://github.com/stnava/ANTsR>). ANTsR is compatible with R versions 3 and above. Directions for compilation can be found in the README.md file associated with the repository. Installation from source assumes one has a compiler and build environment that can be managed with the cross platform build tool CMake which is available at <http://www.cmake.org>. The associated scripts written specifically for the BRATS 2013 challenge are found at <https://github.com/ntustison/BRATS2013>. A simple 2-D example is also provided at this location illustrating both training an RF model and prediction using the previously described framework.

Disclosures The authors declare that they have no conflict of interest.

References

- Amit, Y., & Geman, D. (1997). Shape quantization and recognition with randomized trees. *Neural Computation*, 9, 1545–1588.
- Angelini, E.D., Clatz, O., Mandonnet, E., Konukoglu, E., Capelle, L., Duffau, H. (2007). Glioma dynamics and computational models: A review of segmentation, registration, and in silico growth algorithms and their clinical applications. *Current Medical Imaging Reviews*, 3, 262–276.
- Ashburner, J., & Friston, K. (1997). Multimodal image coregistration and partitioning—a unified framework. *Neuroimage*, 6(3), 209–217. doi:[10.1006/nimg.1997.0290](https://doi.org/10.1006/nimg.1997.0290).
- Avants, B., Duda, J.T., Kim, J., Zhang, H., Pluta, J., Gee, J.C., Whyte, J. (2008a). Multivariate analysis of structural and diffusion imaging in traumatic brain injury. *Academic Radiology*, 15(11), 1360–1375. doi:[10.1016/j.acra.2008.07.007](https://doi.org/10.1016/j.acra.2008.07.007).
- Avants, B.B., Epstein, C.L., Grossman, M., Gee, J.C. (2008b). Symmetric diffeomorphic image registration with cross-correlation: evaluating automated labeling of elderly and neurodegenerative brain. *Medical Image Analysis*, 12(1), 26–41. doi:[10.1016/j.media.2007.06.004](https://doi.org/10.1016/j.media.2007.06.004).
- Avants, B.B., Yushkevich, P., Pluta, J., Minkoff, D., Korczykowski, M., Detre, J., Gee, J.C. (2010). The optimal template effect in hippocampus studies of diseased populations. *Neuroimage*, 49(3), 2457–2466. doi:[10.1016/j.neuroimage.2009.09.062](https://doi.org/10.1016/j.neuroimage.2009.09.062).
- Avants, B.B., Tustison, N.J., Song, G., Cook, P.A., Klein, A., Gee, J.C. (2011a). A reproducible evaluation of ANTs similarity metric performance in brain image registration. *Neuroimage*, 54(3), 2033–2044. doi:[10.1016/j.neuroimage.2010.09.025](https://doi.org/10.1016/j.neuroimage.2010.09.025).
- Avants, B.B., Tustison, N.J., Wu, J., Cook, P.A., Gee, J.C. (2011b). An open source multivariate framework for n-tissue segmentation with evaluation on public data. *Neuroinformatics*, 9(4), 381–400. doi:[10.1007/s12021-011-9109-y](https://doi.org/10.1007/s12021-011-9109-y).
- Bauer, S., Nolte, L.P., Reyes, M. (2011). Fully automatic segmentation of brain tumor images using support vector machine classification in combination with hierarchical conditional random field regularization. *Medical Image Computing and Computer Assisted Intervention*, 14(Pt 3), 354–361.
- Bauer, S., Fejes, T., Slotboom, J., Wiest, R., Nolte, L.P., Reyes, M. (2012). Segmentation of brain tumor images based on integrated hierarchical classification and regularization. In *Proceedings of MICCAI-BRATS 2012*, (pp. 10–13).
- Bauer, S., Wiest, R., Nolte, L.P., Reyes, M. (2013). A survey of MRI-based medical image analysis for brain tumor studies. *Physics in Medicine and Biology*, 58(13), R97–R129. doi:[10.1088/0031-9155/58/13/R97](https://doi.org/10.1088/0031-9155/58/13/R97).
- Breiman, L. (1996). Bagging predictors. *Machine Learning*, 24, 123–140.
- Breiman, L. (2001). Random forests. *Machine Learning*, 45, 5–32.
- Cha, S. (2005). Update on brain tumor imaging. *Current Neurological Neuroscience Report*, 5(3), 169–177.
- Childs, H., Geveci, B., Schroeder, W.J., Meredith, J.S., Moreland, K., Sewell, C., Kuhlen, T., Bethel, E.W. (2013). Research challenges for visualization software. *IEEE Computer*, 46(5), 34–42.
- Criminisi, A., Shotton, J., Konukoglu, E. (2011). *Decision forests for classification, regression, density estimation, manifold learning and semi-supervised learning*. Tech. rep., Microsoft Resaerch.
- Criminisi, A., Robertson, D., Konukoglu, E., Shotton, J., Pathak, S., White, S., Siddiqui, K. (2013). Regression forests for efficient anatomy detection and localization in computed tomography scans. *Medical Image Analysis*. doi:[10.1016/j.media.2013.01.001](https://doi.org/10.1016/j.media.2013.01.001).
- Das, S.R., Avants, B.B., Grossman, M., Gee, J.C. (2009). Registration based cortical thickness measurement. *Neuroimage*, 45(3), 867–879. doi:[10.1016/j.neuroimage.2008.12.016](https://doi.org/10.1016/j.neuroimage.2008.12.016).
- Durst, C.R., Raghavan, P., Shaffrey, M.E., Schiff, D., Lopes, M.B., Sheehan, J.P., Tustison, N.J., Patrie, J.T., Xin, W., Elias, W.J., Liu, K.C., Helm, G.A., Cupino, A., Wintermark, M. (2014). Multimodal MR imaging model to predict tumor infiltration in patients with gliomas. *Neuroradiology*, 56(2), 107–115. doi:[10.1007/s00234-013-1308-9](https://doi.org/10.1007/s00234-013-1308-9).
- Freund, Y., & Schapire, R. (1997). A decision-theoretic generalization of on-line learning and an application to boosting. *Journal of Computer and System Sciences*, 55, 119–139.

- Geremia, E., Clatz, O., Menze, B.H., Konukoglu, E., Criminisi, A., Ayache, N. (2011). Spatial decision forests for MS lesion segmentation in multi-channel magnetic resonance images. *Neuroimage*, 57(2), 378–90. doi:[10.1016/j.neuroimage.2011.03.080](https://doi.org/10.1016/j.neuroimage.2011.03.080).
- Geremia, E., Menze, B.H., Ayache, N. (2012). Spatial decision forests for glioma segmentation in multi-channel MR images. In *Proceedings of MICCAI-BRATS 2012*.
- Gray, K.R., Aljabar, P., Heckemann, R.A., Hammers, A., Rueckert, D. (2013). Alzheimer's Disease Neuroimaging Initiative Random forest-based similarity measures for multi-modal classification of alzheimer's disease. *Neuroimage*, 65, 167–175. doi:[10.1016/j.neuroimage.2012.09.065](https://doi.org/10.1016/j.neuroimage.2012.09.065).
- Greenspan, H.P. (1972). Models for the growth of a solid tumor by diffusion. *Studies in Applied Mathematics*, 52(4), 317–340.
- Ho, T.K. (1995). Random decision forests. In *Proceedings of the 3rd international conference on document analysis and recognition, 1995* (Vol. 1, pp. 278–282). doi:[10.1109/ICDAR.1995.598994](https://doi.org/10.1109/ICDAR.1995.598994).
- Iglesias, J.E., Liu, C.Y., Thompson, P., Tu, Z. (2010). Agreement-based semi-supervised learning for skull stripping. *Medical Image Computing and Computer Assisted Intervention*, 13(Pt3), 147–54.
- Ince, D.C., Hatton, L., Graham-Cumming, J. (2012). The case for open computer programs. *Nature*, 482(7386), 485–488. doi:[10.1038/nature10836](https://doi.org/10.1038/nature10836).
- Kehrer, J., & Hauser, H. (2013). Visualization and visual analysis of multifaceted scientific data: a survey. *IEEE Transaction Vis Computer Graph*, 19(3), 495–513. doi:[10.1109/TVCG.2012.110](https://doi.org/10.1109/TVCG.2012.110).
- Landman, B.A., Huang, A.J., Gifford, A., Vikram, D.S., Lim, I.A.L., Farrell, J.A.D., Bogovic, J.A., Hua, J., Chen, M., Jarso, S., Smith, S.A., Joel, S., Mori, S., Pekar, J.J., Barker, P.B., Prince, J.L., van Zijl, P.C.M. (2011). Multi-parametric neuroimaging reproducibility: a 3-T resource study. *Neuroimage*, 54(4), 2854–2866. doi:[10.1016/j.neuroimage.2010.11.047](https://doi.org/10.1016/j.neuroimage.2010.11.047).
- Liaw, A., & Wiener, M. (2002). Classification and regression by random forest. *R News*, 2(3), 18–22.
- Louis, D.N., Ohgaki, H., Wiestler, O.D., Cavenee, W.K., Burger, P.C., Jouvet, A., Scheithauer, B.W., Kleihues, P. (2007). The 2007 WHO classification of tumours of the central nervous system. *Acta Neuropathologica*, 114(2), 97–109. doi:[10.1007/s00401-007-0243-4](https://doi.org/10.1007/s00401-007-0243-4).
- Maurer, C.R., Rensheng, Q., Raghavan, V. (2003). A linear time algorithm for computing exact Euclidean distance transforms of binary images in arbitrary dimensions. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 25(2), 265–270. doi:[10.1109/TPAMI.2003.1177156](https://doi.org/10.1109/TPAMI.2003.1177156).
- Menze, B., Jakab, A., Bauer, S., Kalpathy-Cramer, J., Farahani, K., Kirby, J., Burren, Y., Porz, N., Slotboom, J., Wiest, R., Lanczi, L., Gerstner, E., Weber, M.A., Arbel, T., Avants, B., Ayache, N., Buendia, P., Collins, L., Cordier, N., Corso, J., Criminisi, A., Das, T., Delingette, H., Demiralp, C., Durst, C., Dojat, M., Doyle, S., Festa, J., Forbes, F., Geremia, E., Glocker, B., Golland, P., Guo, X., Hamamci, A., Iftekharuddin, K., Jena, R., John, N., Konukoglu, E., Lashkari, D., Antonio Mariz, J., Meier, R., Pereira, S., Precup, D., Price, S., Reza, S., Ryan, M., Schwartz, L., Shin, H.C., Shotton, J., Silva, C., Sousa, N., Subbanna, N., Szekely, G., Taylor, T., Thomas, O., Tustison, N., Unal, G., Vasseur, F., Wintermark, M., Hye Ye, D., Zhao, L., Zhao, B., Zikic, D., Prastawa, M., Reyes, M., Van Leemput, K. (2014). The multimodal brain tumor image segmentation benchmark (brats). <http://hal.inria.fr/hal-00935640>.
- Menze, B.H., Van Leemput, K., Lashkari, D., Weber, M.A., Ayache, N., Golland, P. (2010). A generative model for brain tumor segmentation in multi-modal images. *Medical Image Computing Computer Assisted Intervention*, 13(Pt 2), 151–159.
- Nyúl, L.G., Udupa, J.K., Zhang, X. (2000). New variants of a method of MRI scale standardization. *IEEE Transactions on Medical Imaging*, 19(2), 143–150. doi:[10.1109/42.836373](https://doi.org/10.1109/42.836373).
- Padfield, D., & Miller, J. (2008). A label geometry image filter for multiple object measurement. *Insight Journal*, 1–13. <http://hdl.handle.net/1926/1493>.
- Prastawa, M., Bullitt, E., Moon, N., Van Leemput, K., Gerig, G. (2003). Automatic brain tumor segmentation by subject specific modification of atlas priors. *Academic Radiology*, 10(12), 1341–1348.
- Price, S.J., Burnet, N.G., Donovan, T., Green, H.A.L., Pena, A., Antoun, N.M., Pickard, J.D., Carpenter, T.A., Gillard, J.H. (2003). Diffusion tensor imaging of brain tumours at 3T: a potential tool for assessing white matter tract invasion? *Clinical Radiology*, 58(6), 455–462.
- Prima, S., Ourselin, S., Ayache, N. (2002). Computation of the mid-sagittal plane in 3-D brain images. *IEEE Transactions on Medical Imaging*, 21(2), 122–138. doi:[10.1109/42.993131](https://doi.org/10.1109/42.993131).
- Reynolds, D.A. (2009). Gaussian mixture modeling. In S.Z. Li & A.K. Jain (Eds), *Encyclopedia of biometrics* (pp. 659–663). USA: Springer.
- Schapire, R. (1990). The strength of weak learnability. *Machine Learning*, 5, 197–227.
- Tustison, N.J., & Avants, B.B. (2013). Explicit B-spline regularization in diffeomorphic image registration. *Frontiers in Neuroinformatics*, 7, 39. doi:[10.3389/fninf.2013.00039](https://doi.org/10.3389/fninf.2013.00039).
- Tustison, N.J., & Gee, J.C. (2009). Introducing Dice, Jaccard, and other label overlap measures to ITK. *Insight Journal*, 1–4. <http://hdl.handle.net/10380/3141>.
- Tustison, N.J., Avants, B.B., Cook, P.A., Zheng, Y., Egan, A., Yushkevich, P.A., Gee, J.C. (2010). N4ITK: improved N3 bias correction. *IEEE Transactions on Medical Imaging*, 29(6), 1310–1320. doi:[10.1109/TMI.2010.2046908](https://doi.org/10.1109/TMI.2010.2046908).
- Tustison, N.J., Johnson, H.J., Rohlfing, T., Klein, A., Ghosh, S.S., Ibanez, L., Avants, B.B. (2013). Instrumentation bias in the use and evaluation of scientific software: recommendations for reproducible practices in the computational sciences. *Frontiers in Neuroscience*, 7, 162. doi:[10.3389/fnins.2013.00162](https://doi.org/10.3389/fnins.2013.00162).
- VanHorn, J.D., & Toga, A.W. (2013). Human neuroimaging as a “big data” science. *Brain Imaging Behav.* doi:[10.1007/s11682-013-9255-y](https://doi.org/10.1007/s11682-013-9255-y).
- Verhoeck, M., Yaqub, M., McManigle, J., Noble, J.A. (2011). Learning optical flow propagation strategies using random forests for fast segmentation, in dynamic 2D and 3D echocardiographies. In K. Suzuki, F. Wang, D. Shen, P. Yan (Eds.), *Machine learning in medical imaging, lecture notes in computer science* (Vol. 7009, pp. 75–82). Berlin Heidelberg: Springer.
- Viola, P., Jones, M., Snow, D. (2005). Detecting pedestrians using patterns of motion and appearance. *International Journal of Computer Vision*, 63, 153–161.
- Wilkinson, G.N., & Rogers, C.E. (1973). Symbolic description of factorial models for analysis of variance. *Journal of the Royal Statistical Society Series C (Applied Statistics)*, 22(3), 392–399.
- Yi, Z., Criminisi, A., Shotton, J., Blake, A. (2009). Discriminative, semantic segmentation of brain tissue in MR images. *Medical Image Computing Computed Assisted Intervention*, 12(Pt 2), 558–565.
- Zikic, D., Glocker, B., Konukoglu, E., Shotton, J., Criminisi, A., Ye, D.H., Demiralp, C., Thomas, O.M., Das, T., Jena, R., Price, S.J. (2012). Context-sensitive classification forests for segmentation of brain tumor tissues. In *Proceedings of MICCAI-BRATS 2012* (pp. 1–9).